UNCLASSIFIED

AD NUMBER: AD0217057
LIMITATION CHANGES
TO: Approved for public release; distribution is unlimited.
FROM: Distribution authorized to U.S. Gov't. agencies and their contractors; Administrative/Operational Use; 1 Apr 1959. Other requests shall be referred to Army Quartermaster Research Engineering Command, Natick MA 01760
AUTHORITY
USAETL LTR 28 SEP 1976

THIS PAGE IS UNCLASSIFIED

UNCLASSIFIED

AD 217057

Armed Services Technical Information Agency

ARLINGTON HALL STATION
ARLINGTON 12 VIRGINIA

FOR
MICRO-CARD
CONTROL ONLY

1 OF 1

UNCLASSIFIED

HEADQUARTERS
QUARTERMASTER RESEARCH & ENGINEERING COMMAND
U S ARMY

TECHNICAL REPORT
EP-112

PREDICTIVE METHODS IN TOPOGRAPHIC ANALYSIS
I. RELIEF, SLOPE, AND DISSECTION ON INCH-TO-THE-MILE MAPS IN THE UNITED STATES

QUARTERMASTER RESEARCH & ENGINEERING CENTER
ENVIRONMENTAL PROTECTION RESEARCH DIVISION

APRIL 1959 NATICK, MASSACHUSETTS

HEADQUARTERS
QUARTERMASTER RESEARCH & ENGINEERING COMMAND, US ARMY
OFFICE OF THE COMMANDING GENERAL
NATICK, MASSACHUSETTS

Major General Andrew T. McNamara
The Quartermaster General
Washington 25, D. C.

Dear General McNamara:

This report, "Predictive Methods in Topographic Analysis: I. Relief, Slope, and Dissection on Inch-to-the-Mile Maps in the United States," is the first of a series which will supply the Quartermaster Corps and the Army with quantitative terrain information and with methods for predicting the topographic characteristics of small areas of the earth's surface. This research will result in a numerical system for describing the geometric dimensions of parts of the earth's surface. It will also supply methods for producing the information required by the system.

At the present time, descriptions of terrain are expressed in qualitative terms which everyone interprets in the light of his own experience. This report demonstrates that it is possible to describe terrain with numbers which have the same meaning for everyone. Quantitative descriptions of terrain can supply information needed in the writing of criteria for designing specific equipment (for example, vehicles), and in the formulation of logistical and tactical plans.

Sincerely yours,

C G Calloway

C. G. CALLOWAY
Major General, USA
Commanding

1 Incl:
EP-112

HEADQUARTERS
QUARTERMASTER RESEARCH & ENGINEERING COMMAND, US ARMY
Quartermaster Research & Engineering Center
Natick, Massachusetts

ENVIRONMENTAL PROTECTION RESEARCH DIVISION

Technical Report
EP-112

PREDICTIVE METHODS IN TOPOGRAPHIC ANALYSIS

I: RELIEF, SLOPE, AND DISSECTION ON INCH-TO-THE-MILE MAPS IN THE UNITED STATES

Walter F. Wood, Ph.D.
Geographer

Joan B. Snell, M.A.
Geographer

Environmental Analysis Branch

Project Reference:
7-83-01-005

April 1959

Foreword

Precise numerical definitions of terrain characteristics are needed in order that terrain information may be mathematically applied to such problems as the design of military equipment and logistical systems. The present terrain analysis study has been undertaken to attain this objective.

Prior to establishing a numerical system for describing landscapes, the fundamental relationships existing among terrain elements must be understood. Also, methods must be developed for obtaining basic terrain data quickly, not only from existing maps but for areas where topographic mapping is not available. A series of reports is proposed to furnish such information.

This report, the first of the series, describes the relationships of six geomorphic factors and introduces formulae for predicting three of them.

AUSTIN HENSCHEL, Ph.D.
Chief
Environmental Protection Research Division

Approved:

CARL L. WHITNEY, Lt Col, QMC
Commanding Officer
QM R and E Center Laboratories

J. FRED OESTERLING, Ph.D.
Acting Scientific Director
QM Research & Engineering Command

Contents

Abstract

Based on a 200-case sample of United States topographic maps, a study ~~has been~~ was made of the inter-relationships of six terrain factors: highest elevation, lowest elevation, relief, contour counts, slope direction changes, and hilltops. The rank method of correlation was used for this analysis. From the preliminary correlations, three factors (relief, contour counts, and slope direction changes) were found to be most significantly related. Further work with these three factors resulted in a series of regression equations, both simple and multiple, to be used for predicting any of these three elements when pertinent data for it are lacking.

The performance of two multiple regression equations to predict average slope from relief and from slope direction changes, were mapped. It was seen that average slope can be predicted within an error of 4% slope for three-fourths of the United States. The best predictions were made for the Coastal Plain and Interior Lowlands, areas of low relief. The poorest predictions were in the Appalachians, parts of the Northern Rockies, Cascades, Colorado Plateau, and Ozarks.

PREDICTIVE METHODS IN TOPOGRAPHIC ANALYSIS: I. RELIEF, SLOPE, AND DISSECTION ON INCH-TO-THE-MILE MAPS IN THE UNITED STATES

1. Purpose and scope

The Army operates on the earth's surface and therefore considers the description of this surface to be of vital concern. Heretofore, terrain has been described in qualitative terms - steep, rough, rolling, flat - which are subject to misinterpretation and are inadequate for many purposes. There is a need, then, for explicit description of terrain. Since the surface of the earth can be measured, quantitative descriptions of this surface are possible and will provide much needed information.

Quantified terrain data can be applied to the development of criteria for the design and testing of military equipment. Also, this knowledge can provide a uniform method for evaluating field performance of equipment systems with respect to terrain. Moreover, quantitative descriptions of such factors as "relief" and "average slope" will be useful in other studies of the environment such as meteorology and trafficability. It is known, for instance, that the roughness of the earth's surface affects the speed and direction of wind and that there is a relationship between size of soil particle and angle of slope.

To fulfill these many needs, a numerical system for describing landscapes should be devised. This system should be simple, yet complete enough to carry all pertinent information. Adaptability to machine processing of these data, along with those on military equipment or other environmental data, is also desirable. Such a simple and meaningful system, or code, must be based upon a knowledge of the interrelationship among individual characteristics of the earth's surface. The task of supplying terrain information and designing a code would be relatively simple if the earth were mapped entirely at an inch-to-the-mile scale with a 20-foot contour interval. Unfortunately, the only maps available for much of the world are at a scale of 1:1,000,000 (15.8 miles to the inch) with a 1,000-foot contour interval. Thus, the lack of detailed topographic maps complicates the task of collecting basic terrain data.

The research here described was undertaken to discover interrelationships among terrain elements and to demonstrate how prediction equations, which will help to overcome the limitations of inadequate map coverage, may be constructed, applied, and tested. In their present state of development, the prediction methods have some military utility and, as more work is done, the applicability of the findings will become more precise.

This report, the first of a series presenting the results of a continuing terrain analysis study, will deal only with interrelationships of terrain factors and prediction methods, but the ultimate objective of a code was in mind, as the research proceeded. The paper is further limited to a consideration of those topographic characteristics which can be collectively called surface geometry.

2. Previous quantitative investigations

That order exists on the earth's surface was first demonstrated by John Playfair, a professor of mathematics at the University of Edinburgh. In 1802, Playfair published a commentary on Hutton's "Theory of the Earth" in which he (Playfair) stated: "Every river appears to consist of a main trunk, fed from a variety of branches, each running in a valley proportioned to its size, and all of them together forming a system of vallies (sic), communicating with one another, and having such a nice adjustment of their declivities that none of them join the principal valley, either on too high or too low a level."(6) This observation is known as "Playfair's Law" of concordant stream junctions and, although expressed qualitatively, it has been the basis for recent hydrological investigations.

A quantitative interpretation of Playfair's Law was undertaken by the hydrologist, R. E. Horton. In 1945, Horton showed that stream channels develop according to definite hydrophysical laws. Thus he was able to express the composition of the stream system of a drainage basin quantitatively in terms of stream order, drainage density, bifurcation ratio, and stream-length ratio.(3) Since that time, other hydrologists have elaborated upon and refined the basic concepts which Horton described.

These first quantitative studies were concerned with stream systems. However, if stream systems have order, their interfluves, which have been shaped by them, must also exhibit order. Recently geologists and geographers have turned to a quantitative interpretation of other elements of the earth's surface. In 1950, Strahler used statistical methods to test alternative hypotheses in geology.(7) Other studies by Strahler and his associates* have made it abundantly clear that the apparent disorder, which had previously been seen in the geometry of the earth's surface, is, in reality, order.

From the above statements it should follow that randomly chosen units of the earth's surface should have mathematical relationships among their separate dimensions. Also, it is likely that information about large areas gathered from small-scale maps, would have a relationship to the smaller areas contained within them. If this is true, the possibility exists that by using regression relationships, predictions can be made about small areas which have not been mapped at topographic scale.

* Working under Office of Naval Research Contract N6 ONR 271.

So far as is known, only three probing attempts have been made to discover whether it is possible to develop regression equations for predicting terrain characteristics on randomly chosen parts of the earth's surface. Peltier, in 1954, found that average slope is empirically related to, and varies directly with, the average local relief from one-square-mile areas.[5] The following year, another investigation showed that regression equations for predicting the relief on a square mile became increasingly more accurate as a set of terrain samples was divided and subdivided on the basis of natural vegetation type and geologic bedrock.[10] A third study, one towards developing prediction equations, was made in 1956 when it was demonstrated that average slope on rather small areas (9 square kilometers) could be predicted through the use of data gathered from maps with scales of 1:250,000 and 1:1,000,000.[9] These three studies were small in terms of the amount of data involved and their only purpose was to investigate the possibility of developing statistical methods for predicting relief or average slope of randomly chosen small areas.

The results of the works cited above were promising, and a much more ambitious project, involving more data and much more extensive analysis was undertaken by the present authors. Means, modes, medians, and standard deviations of the terrain factors to be treated in this latest study have already been published.[11]

3. Present quantitative study

a. Basic data

The sample selected for study consists of a group of 204 US topographic maps chosen at random from the Army Map Service index; these maps are at a scale of 1:62,500 or 1:50,000. These sheets represent every major physiographic province of the United States and include a wide variety of terrain features (Fig. 1).

From this sample, data were gathered for six terrain elements. These were:

By area:

1. Highest elevation within a sample unit.
2. Lowest elevation within a sample unit.
3. Relief (difference between the highest and lowest elevations of the sample unit) within a sample unit.
4. Number of hilltops (individual heights of land as represented by a closed contour) within a sample unit.

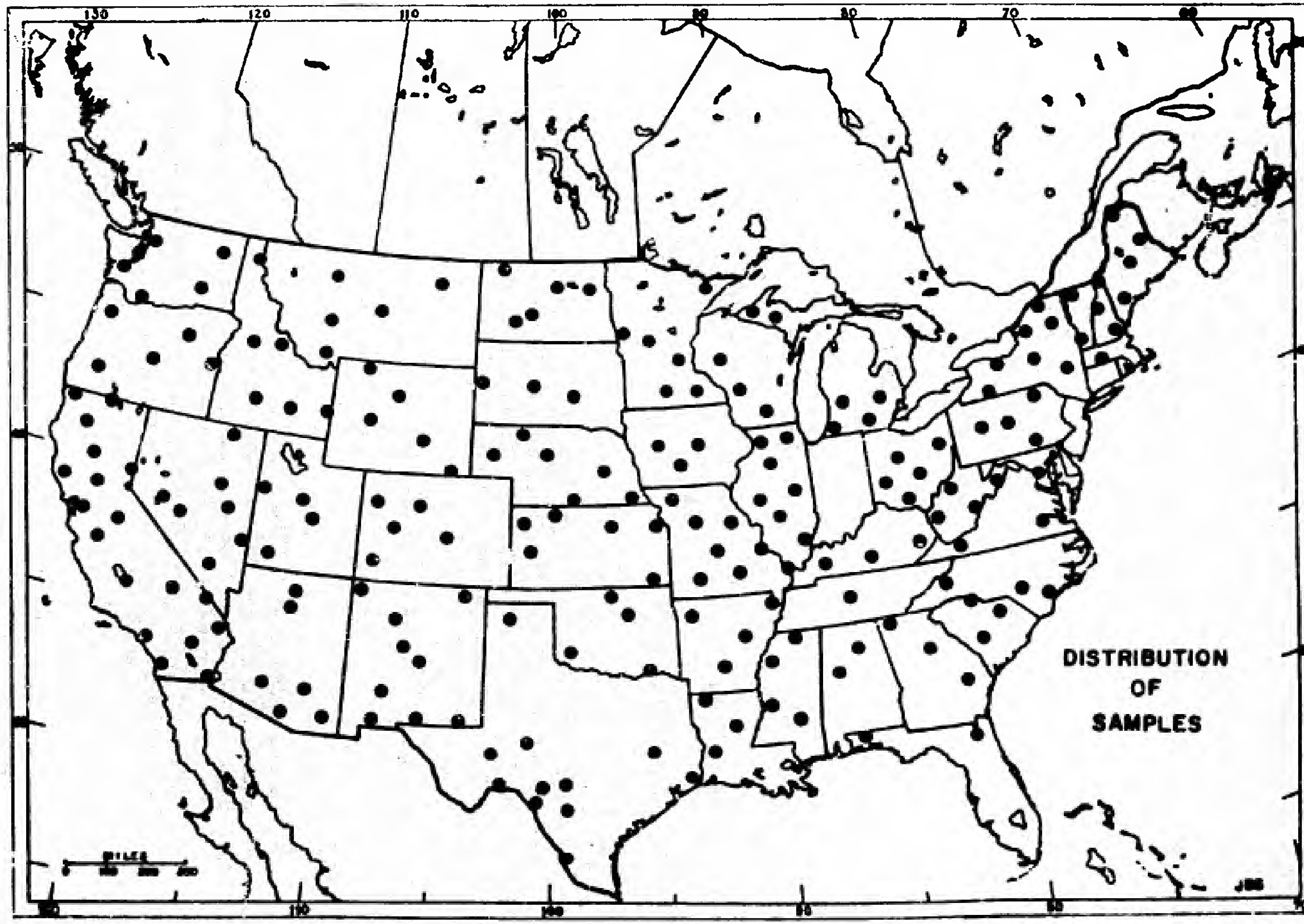


Figure 1. Distribution of the 204 topographic maps of the United States which provide a representative terrain sample.

By traverse:

5. Number of 20-foot contour crossings (contour counts) along a traverse within the sample unit, used as a measure of average slope.
6. Number of valleys and divides encountered on the same traverse, used as a measure of dissection and collectively referred to as slope direction changes.

The smallest sample unit is a circle with an area of 5/16 of a square mile, the mid-point of which is the center of the topographic sheet. Successive doublings of this unit produced sample circles with areas of 5/8, 1¼, 2½, 5, 10, 20, 40, 80, and 160* square miles, resulting in 10 sample sizes. Two diameters were drawn in the cardinal directions through the mid-point of the concentric circles. These lines were used as traverses along which contours and slope direction changes were counted. The combined lengths of the two diameters for each circle are 1.2, 1.8, 2.5, 3.6, 5.1, 7.1, 10.1, 14.3, 20.2 and 28.5 miles (Fig. 2).

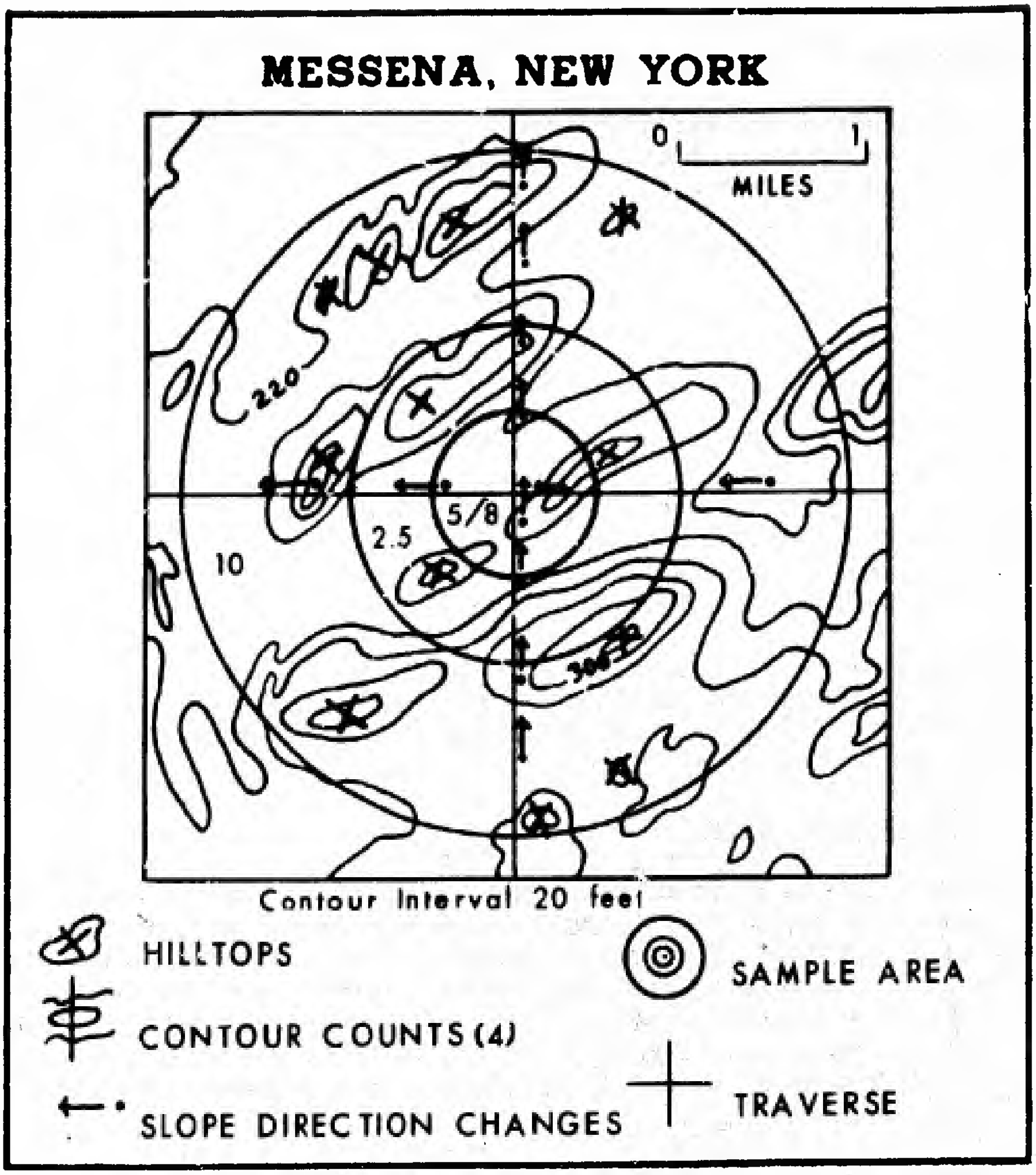
Thus there were ten area sizes per sheet, with a total of 2,040 individual sample units to be considered. For each sample unit there were gathered the previously enumerated information on the six terrain elements, making a total of 12,240 pieces of data for analysis.

b. Preliminary analysis

The first analysis of these data revealed an orderly progression in the change of value of individual measures with change in size of area. The data also exhibited a strong skewness which all attempts failed to normalize. However, if several statistical distributions refuse to be normalized, they can be transformed into identical rectangular distributions prior to analysis. This will assure that extreme values within the population do not exert an undue influence on any regression equation which may be developed.

A quick analysis of twenty cases drawn from the total samples indicated the feasibility of using rank correlation methods for ease of interpreting data. The relief values of these cases were ranked for each size of area, i.e., arranged in ascending order of magnitude and assigned a value from 1 to 20, with the lowest value of relief becoming 1 and the highest becoming 20. These rank values for each size of area were then correlated with those of every other size area. The coefficients of rank correlation showed an orderly change of value with change in area of samples, thus indicating the degree of success to be expected if relief of a large area is used to predict relief of a small area which lies at its center.

* A few of the 160-square-mile samples had to be arranged as elipses to stay within the limits of the topographic map.



	5/8 sq. mi.	2.5 sq. mi.	10 sq. mi.
Highest Elevation - feet	260	300	320
Lowest Elevation - feet	220	220	220
Relief - feet	40	80	100
Hilltops	0	3	12
Contour Counts	8	14	29
Slope Direction Changes	4	6	12

Figure 2. Portion of U.S.G.S. topographic map illustrating data-gathering technique. Highest and lowest elevations read to the nearest contour. When using a topographic map for the gathering of data, interpretations are made for elevations both above highest contour and below lowest contour.

It was also shown that the values of the coefficients of correlation increased in an orderly fashion as the size of the independent variables approached the size of the dependent variable. This orderliness was so pronounced that for the purpose of future study, which would be done by rank correlation methods, it was deemed safe to eliminate half of the sample according to size of areas and concentrate only on the 5/8, 2½, 10, 40, and 160-square-mile samples.

4. Relationship among terrain elements

a. Correlation by the rank method

Rank correlation does not offer the only possible method for analyzing the data. Graphical curvilinear methods, for instance, would be applicable, and in some cases give more accurate results. However, such an approach is laborious and, when more than one independent variable is used, extremely slow. By adopting the rank correlation method, it is possible to put the data on punch cards and do much of the computation on machines, thereby greatly shortening the task. Strictly speaking, for this work, the Pearsonian equation for correlation $r_r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}}$ was used with rank values substituted for absolute values instead of the usual equation for computing rank correlations $r_r = 1 - \frac{6 \Sigma d^2}{N(N^2 - 1)}$. The Pearsonian formula is more readily applied to the computations furnished by the EAM machines. When 200 paired values are used for a correlation, the results are so nearly the same that accuracy is not sacrificed. (Rank correlation of contour counts and slope direction changes on 5/8 square miles: Pearsonian formula: $r = .5193$; rank formula: $r_r = .5228$.)

Simple correlations reveal the degree of interrelationship among individual factors and indicate which ones can be used to predict others. The degree of interrelationship is indicated by the end product of correlations, the value r_r. This measure notes the significance between elements and can be evaluated from the charts or tables of reliability found in any standard statistics book.

For this study of 200 cases (for ease of statistical manipulation, four of the original 204 samples were discarded) an r_r value of .1235 is required for the 5% level of significance*, according to the F table of significance.[(1)] However, for one factor to be a good predictor of another, the r_r value must be considerably higher. A useful indicator of prediction accuracy is the r_r^2 which will tell what percentage of total variance between two elements has been accounted for by the correlation. Thus an r_r value of .1235, although significant in that it probably did not occur by chance, accounts for only 2% of the total variance

* This value of r_r would occur only once in 20 times by chance.

TABLE I

COEFFICIENTS OF RANK CORRELATION (r_r)
(for six terrain factors within sample areas)

	Highest Elevation	Lowest Elevation	Relief	Hilltops	Contour Counts	Slope Direction Changes
			5/8 Square Miles			
High Elevation	----	.9799	.5507	.0526*	.5054	.0782*
Low Elevation		----	.4247	.0386*	.3863	.0439*
Relief			----	.2213	.9472	.3957
Hilltops				----	.2935	.4078
Contour Counts					----	.5193
Slope Dir Changes						----
			2 1/2 Square Miles			
High Elevation	----	.9619	.6277	.0716*	.5235	.1106*
Low Elevation		----	.4526	.0538*	.3604	.0677*
Relief			----	.2395	.9291	.3752
Hilltops				----	.3559	.5268
Contour Counts					----	.5567
Slope Dir Changes						----
			10 Square Miles			
High Elevation	----	.9402	.7009	.0734*	.5357	.0305*
Low Elevation		----	.4755	.0388*	.3176	-.0299*
Relief			----	.1967	.8911	.2684
Hilltops				----	.3715	.5327
Contour Counts					----	.5332
Slope Dir Changes						----
			40 Square Miles			
High Elevation	----	.9086	.7707	.1036*	.5712	-.0238*
Low Elevation		----	.4969	.0656*	.3148	-.0466*
Relief			----	.2025	.8632	.1549
Hilltops				----	.3471	.5006
Contour Counts					----	.4651
Slope Dir Changes						----
			160 Square Miles			
High Elevation	----	.8619	.8464	.1381#	.6079	-.0606*
Low Elevation		----	.5215	.1071*	.3143	-.0555*
Relief			----	.1935	.8287	.0671*
Hilltops				----	.3029	.4239
Contour Counts					----	.4153
Slope Dir Changes						----

* not significant r_r at 5% level (r_r = .1235 required for significance)
\# not significant r_r at 1% level (r_r = .1525 required for significance)

9

TABLE II

COEFFICIENTS OF RANK CORRELATION (r_r)
(for three terrain factors[1] within and between sample areas)

Square Miles		5/8 Square Miles			2-1/2 Square Miles			10 Square Miles			40 Square Miles			160 Square Miles		
		Rel.	C.C.	S.D.C.	Rel.	C.C.	S.D.C.	Rel.	C.C.	S.D.C.	Rel.	C.C.	S.D.C.	Rel.	C.C.	S.D.C.
5/8	Rel.	----	.9472	.3957	.97	.9529	.4364	.9044	.9438	.4174	.8493	.9199	.3658	.7856	.8848	.3079
	C.C.		----	.5193	.9121	.9851	.5563	.8510	.9602	.5281	.7886	.9326	.4769	.7245	.8955	.4141
	S.D.C.			----	.3430	.4959	.8907	.2742	.4595	.7907	.1966	.4144	.7348	.1344#	.3868	.6824
2 1/2	Rel.				----	.9291	.3752	.9547	.9422	.3622	.9019	.9358	.3222	.8384	.9110	.2789
	C.C.					----	.5567	.8643	.9841	.5385	.7966	.9558	.4695	.7295	.9148	.4217
	S.D.C.						----	.2990	.5249	.9162	.2186	.4717	.8393	.1509#	.4305	.7599
10	Rel							----	.8911	.2684	.9659	.9142	.2440	.7182	.9142	.2168
	C.C.								----	.5332	.8255	.9809	.4962	.7515	.9448	.4431
	S.D.C.									----	.1727	.4751	.9332	.1021*	.4250	.8625
40	Rel.										----	.8632	.1599	.9636	.8873	.1418#
	C.C.											----	.4651	.7[illegible]2	.9798	.4275
	S.D.C.												----	.0[illegible]4*	.4209	.9451
160	Rel													----	.8287	.0671*
	C.C.														----	.4153
	S.D.C.															----

(1) The three terrain factors: Relief (Rel.); Contour Count (C.C.); Slope Direction Change (S.D.C.)
* Not significant r_r at 5% level (r_r = .1235 required for significance)
\# Not significant r_r at 1% level (r_r = .1525 required for significance)

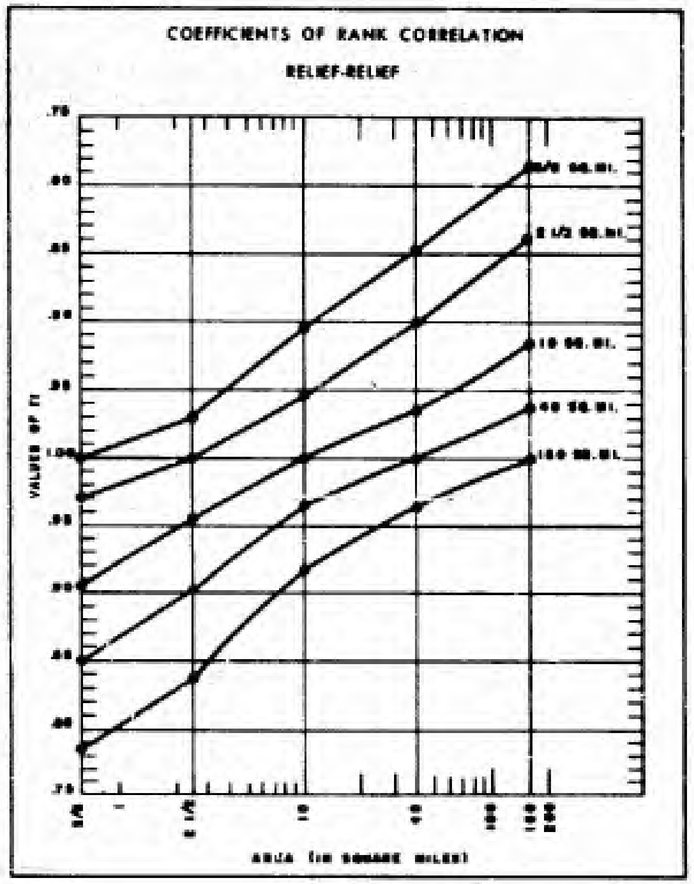


Figure 3. Graph of coefficients of rank correlation: relief to relief on five sizes of area. Correlation coefficients for which data are lacking can be interpolated on other sizes of area between 5/8 and 160 square miles.

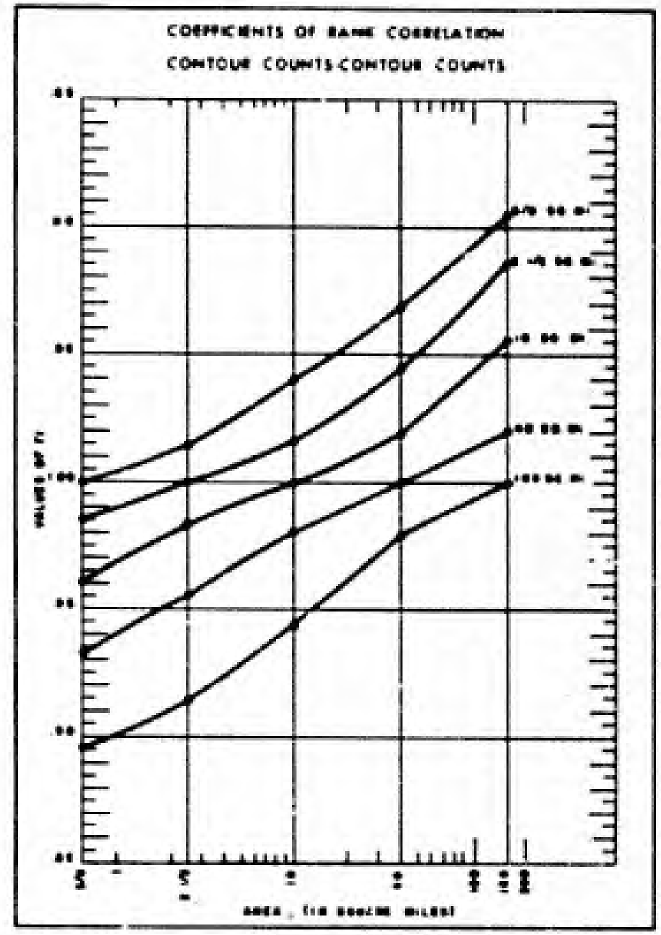


Figure 4. Graph of coefficients of rank correlation: contour counts to contour counts on five sizes of area. Correlation coefficients for which data are lacking can be interpolated on other sizes of area between 5/8 and 160 square miles.

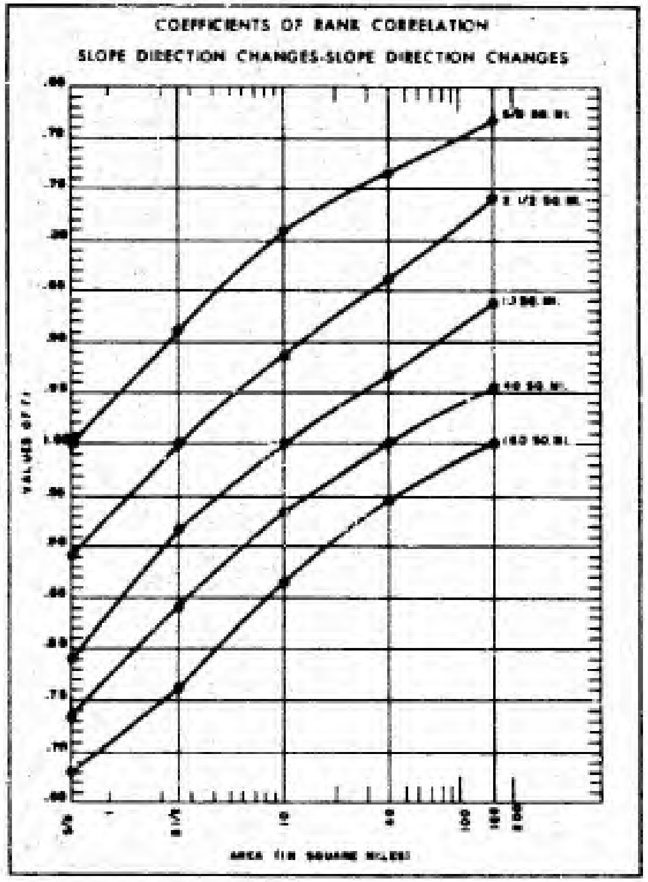


Figure 5. Graph of coefficients of rank correlation: slope direction changes to slope direction changes on five sizes of area. Correlation coefficients for which data are lacking can be interpolated on other sizes of area between 5/8 and 160 square miles.

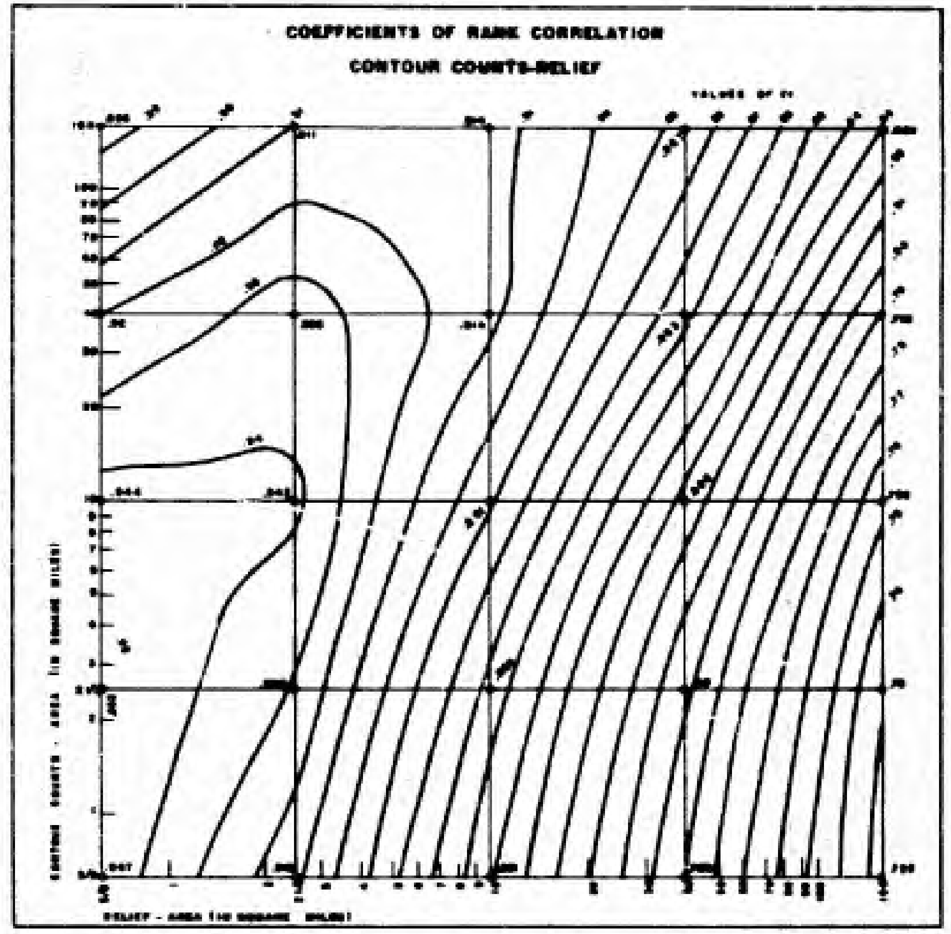


Figure 6. Graph of coefficients of rank correlation: contour counts to relief on five sizes of area. Correlation coefficients for which data are lacking can be interpolated on other sizes of area between 5/8 and 160 square miles.

($.1235^2 = .0152$) and hence would be a poor predictor. On the other hand, an r_r of .92 accounts for about 85% of the total variance ($.92^2 = .8464$). It is of academic interest to know that all terrain elements are related, but only those of relatively <u>high</u> correlation can be used for prediction purposes.

When each terrain characteristic is correlated with each of the other five characteristics within each sample size, 54 of the 75 correlations are found to be significant (Table I).

A close study of Table I, as well as other considerations, permitted the elimination of highest elevation, lowest elevation, and hilltops from further study. Highest elevation and lowest elevation have high correlations only with each other and with relief. (Elevations correlate well with relief, since, by definition, relief is the difference between highest elevation and lowest elevation.) Although "hilltops" (defined as any summit represented by at least one closed contour) is an important element of the landscape, this too was eliminated from further study. This characteristic of terrain is very sensitive to differences in contour interval and map quality. It is expected that rigid control of these two variables in a sample of topographic maps might well bring to light relationships of hilltops to other aspects of surface geometry.

Relief, slope direction changes, and contour counts not only correlate well with each other when sample size remains constant but the relationships often remain close even when the sample sizes being considered are widely different (Table II). From this table it is noted that the r_r values of the various terrain elements may change relatively little when correlated with the same terrain element on another size of area. For instance, an r_r value of .9472 of relief and contour counts on 5/8 square miles, decreases to a value of .8848 when relief on 5/8 square miles is correlated with contour counts on 160 square miles. Nevertheless, the correlations remain high and can be used for predicting, with the exception of relief and slope direction changes.

The change in the values of the coefficients of correlation with change in size of area is so regular as to make possible the construction of graphs which show what the correlations would probably have been, had sizes of area other than the five chosen been used. (Coefficients of rank correlation, in relation to area, have been plotted for relief, contour counts, and slope direction changes, Figures 3, 4, 5, 6.) For example, one might want to know how well relief on 10 square miles correlates with relief on 80 square miles. From Figure 3 select the line on the bottom of the chart which represents 80 square miles. Follow this line upward to the curve which represents 10 square miles. At this point, the value of r_r will be determined. The correlation coefficient of relief on 10 square miles to relief on 80 square miles

is about .94. The other graphs (Figs. 4, 5, and 6) will supply interpolated information regarding contour counts and slope direction changes in the same manner.

The correlation coefficients for the three factors on all sizes of area supply the basic information needed prior to the making of prediction formulae. The tables of r_r (Tables I and II) indicate the significance of one factor to another. The numerical values of r_r can be used for computing simple regression equations based on rank correlations.

b. Predictions

(1) Simple regression equations. Regression equations were computed to predict rank values for the following combinations of terrain factors on all sizes of area:

	Table
1. Relief to relief	III
2. Contour counts to contour counts	IV
3. Slope direction changes to slope direction changes	V
4. Relief to contour counts	VI
5. Slope direction changes to contour counts	VII

Relief to slope direction changes has been omitted because the correlation values indicated that not much of the variance in one factor can be attributed to differences in the other.

The standard procedure of the least squares method of fit was used to estimate the regression, $T = a + b X$, "T" is the prediction, "a" a constant, "b" the coefficient of regression, and "X" the independent variable. With rank correlations, the computations for this formula are not complex, since "b" is identical to the r_r values, and "a" is computed from the formula: $a = \frac{(N + 1)(1 - r_r)}{2}$ where "N" is the number of cases, and "r_r" is the coefficient of rank correlation.

The sample size for this study was 200, hence the regression formula for any factor in this study becomes: $T = \frac{(201)(1 - r_r)}{2} + r_r X$. Should one wish to devise regression equations from information taken from the graphs of r_r values (Figs. 3, 4, 5, and 6), this formula would be used.

TABLE III

SIMPLE REGRESSION EQUATIONS FOR PREDICTING RELIEF FROM RELIEF
(Of Various Size Areas Sharing a Common Center)

Relief

Relief	5/8 Square Mile	2-1/2 Square Miles	10 Square Miles	40 Square Miles	160 Square Miles
5/8 sq mi		T = 3.015 + .97X	T = 9.6078 + .9044X	T = 15.1454 + .8493X	T = 21.5472 + .7896X
2-1/2 sq mi			T = 4.5527 + .9547X	T = 9.8591 + .9019X	T = 16.2408 + .8384X
10 sq mi				T = 3.4271 + .9659X	T = 18.2709 + .9182X
40 sq mi					T = 3.6582 + .9636X
160 sq mi					

TABLE IV

SIMPLE REGRESSION EQUATIONS FOR PREDICTING CONTOUR COUNTS FROM CONTOUR COUNTS
(Of Various Size Areas Sharing a Common Center)

Contour Counts

Contour Counts	5/8 Square Mile	2-1/2 Square Miles	10 Square Miles	40 Square Miles	160 Square Miles
5/8 sq mi		T = 1.4975 + .9851X	T = 3.9999 + .9602X	T = 6.7737 + .9326X	T = 10.5023 + .8955X
2-1/2 sq mi			T = 1.598 + .9841X	T = 4.4421 + .9558X	T = 8.5626 + .9148X
10 sq mi				T = 1.92 + .9809X	T = 5.5476 + .9448X
40 sq mi					T = 2.0904 + .9792X
160 sq mi					

TABLE V

SIMPLE REGRESSION EQUATIONS FOR PREDICTING SLOPE DIRECTION CHANGES FROM SLOPE DIRECTION CHANGES
(Of Various Size Areas Sharing a Common Center)

Slope Direction Changes

Slope Direction Changes	5/8 Square Mile	2/1/2 Square Miles	10 Square Miles	40 Square Miles	160 Square Miles
5/8 sq mi		T = 10.9847 + .8907X	T = 21.0347 + .7907X	T = 26.6526 + .7348X	T = 31.9188 + .6824X
2-1/2 sq mi			T = 8.4219 + .9162X	T = 16.1504 + .8393X	T = 24.13 + .7599X
10 sq mi				T = 6.7134 + .9332X	T = 13.8188 + .8625X
40 sq mi					T = 5.5175 + .9451X
160 sq mi					

TABLE VI

SIMPLE REGRESSION EQUATIONS FOR PREDICTING RELIEF FROM CONTOUR COUNTS OR CONTOUR COUNTS FROM RELIEF

(Of Various Size Areas Sharing a Common Center)

Contour Counts

		5/8 Square Mile	2-1/2 Square Miles	10 Square Miles	40 Square Miles	160 Square Miles
Relief	5/8 sq mi	T = 5.3064 + .9472X	T = 4.7336 + .9529X	T = 5.6481 + .9438X	T = 8.05 + .9195X	T = 11.5776 + .8848X
	2-1/2 sq mi	T = 8.834 + .9121X	T = 7.1255 + .9291X	T = 5.8089 + .9422X	T = 6.4521 + .9358X	T = 8.9445 + .911 X
	10 sq mi	T = 14.9745 + .852 X	T = 13.6379 + .8643X	T = 10.9445 + .8911X	T = 8.6229 + .9142X	T = 8.6229 + .9142X
	40 sq mi	T = 21.2457 + .7886X	T = 20.4417 + .7966X	T = 17.5373 + .8255X	T = 13.7484 + .8632X	T = 11.3264 + .8873X
	160 sq mi	T = 27.6878 + .7245X	T = 27.1853 + .7295X	T = 24.9743 + .7515X	T = 21.2859 + .7882X	T = 17.2157 + .8807X

TABLE VII

SIMPLE REGRESSION EQUATIONS FOR PREDICTING SLOPE DIRECTION CHANGES FROM CONTOUR COUNTS OR CONTOUR COUNTS FROM SLOPE DIRECTION CHANGES

(Of Various size Areas Sharing a Common Center)

Contour Counts

		5/8 Square Mile	2-1/2 Square Miles	10 Square Miles	40 Square Miles	160 Square Miles
Slope Direction Changes	5/8 sq mi	T = 48.3104 + .5193X	T = 50.662 + .4959X	T = 54.3203 + .4595X	T = 58.8528 + .4144X	T = 61.6866 + .3868X
	2-1/2 sq mi	T = 44.5919 + .5563X	T = 44.5517 + .5567X	T = 47.7476 + .5249X	T = 53.0942 + .4717X	T = 57.2348 + .4305X
	10 sq mi	T = 47.426 + .5281X	T = 46.3808 + .5385X	T = 46.9134 + .5339X	T = 52.7525 + .4751X	T = 57.7875 + .425 X
	40 sq mi	T = 52.5726 + .4769X	T = 53.3153 + .4695X	T = 50.6319 + .4962X	T = 53.7575 + .4651X	T = 58.1996 + .4209X
	160 sq mi	T = 58.883 + .4141X	T = 58.1192 + .4217X	T = 55.9685 + .4431X	T = 57.5363 + .4275X	T = 58.7624 + .4153X

T = Prediction - expressed as a rank value
X = Independent Variable - expressed as a rank value

NOTE: T and X in all equations (Tables III, IV, V, VI, and VII) are reversible: i.e., from Table VI - to predict Relief on 5/8 sq. mi. from Contour Counts on 5/8 sq. mi.: T (Rel. 5/8) = 5.3064 + .9472X (C.C. 5/8). Or, to predict Contour Counts on 5/8 sq. mi. from Relief on 5/8 sq. mi.: T (C.C. 5/8) = 5.3064 + .9472X (Rel. 5/8).

An example of the application of these formulae follows. Data are from Goldhill, Utah, USGS topographic map.(2) The relief of a 10-square-mile area is known (1,691 feet) and one wants to determine the relief on 5/8 square miles within that area. The equation $T = 9.6078 + .9044X$ will be used, where "T" is the relief of 5/8 square miles and "X" is the relief of 10 square miles (Table III). The relief value on 10 square miles, 1,691 feet is converted to a rank value, 173. (See Nomograph to convert actual values of relief to rank values of relief, Figure 7. Nomographs for converting actual values to rank values have also been included for contour counts and slope direction changes, Figures 8 and 9.) Substitutions are made in the equation and the equation solved:

$$T = 9.6078 + (.9044 \times 173)$$
$$T = 9.6078 + 156.46$$
$$T = 166.06$$

resulting in a prediction of relief for the 5/8-square-mile area as rank 166.06. Predicted rank value 166.06 is converted to a predicted actual value for relief: 580 feet, by referring to the nomograph (Fig. 7). The relief of this sample is 560 feet, so the prediction is only 20 feet more than what actually occurs.

Simple regression equations to predict relief, slope direction changes, and contour counts have been developed for the area sizes studied and it has been shown that additional equations of this type could be developed. With these equations one should be able to make fairly accurate estimates of these elements on any size of area when information for only one of the factors is available.

(2) Multiple regression equations

In the preceding section, relief, contour counts, and slope direction changes have been built individually into simple regression equations for predicting various combinations of these factors. When a minimum of terrain information is available, the best predictions that can be hoped for are obtained by using these equations. With additional information, however, the original predictions can be improved through the development of multiple regression equations.

Multiple regression is similar to simple regression, but makes use of more than one factor for predicting another. Theoretically, it would be possible to determine all the important elements that constitute an individual piece of the land surface. All these elements except one could then be built into a multiple regression equation which would, theoretically at least, give near perfect estimates of the missing one. However, at this stage of terrain research, all the contributing factors are not known; neither is adequate information available.

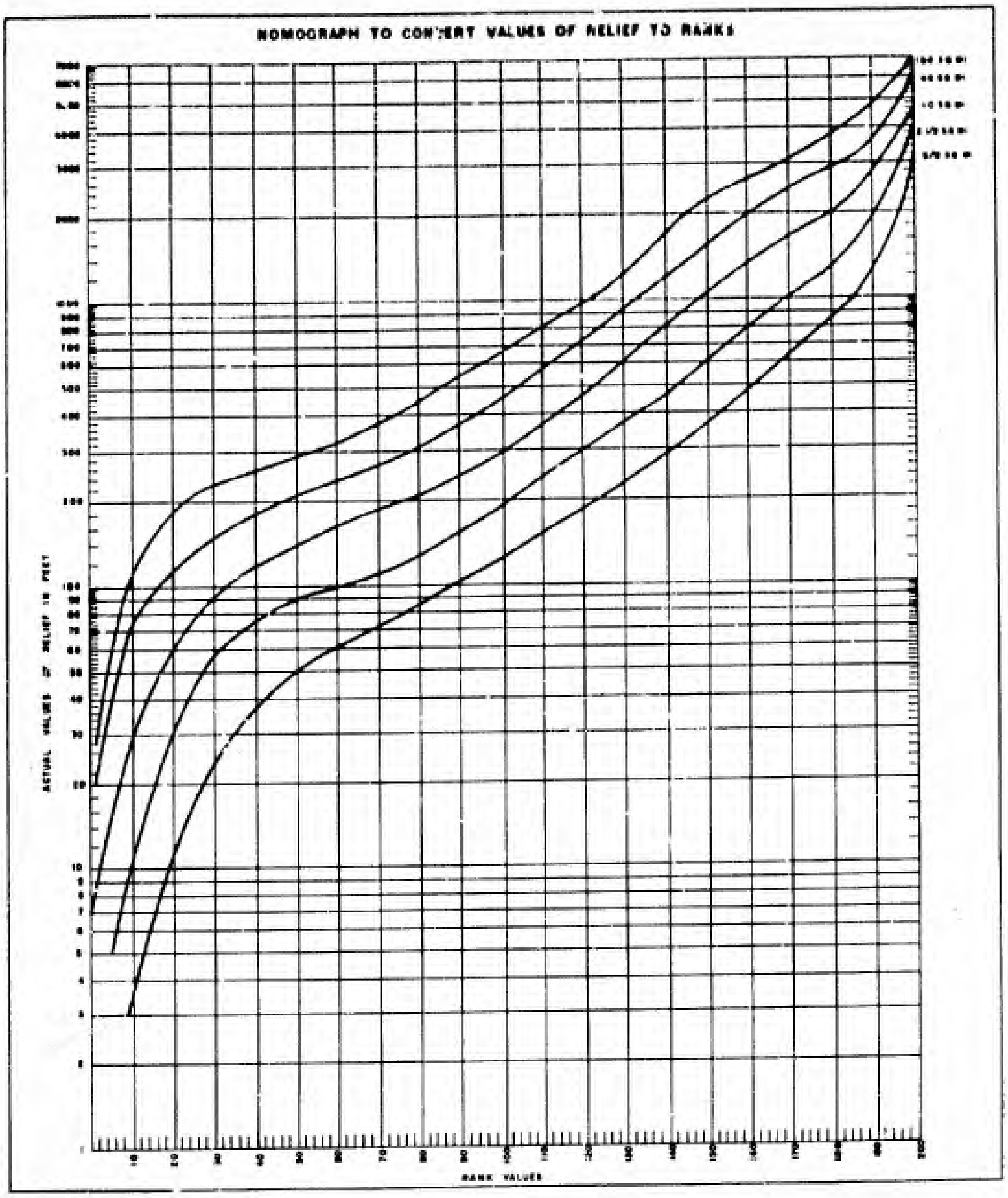


Figure 7. For use in the prediction equations, actual values of relief on five sizes of area are converted to rank values of relief, or the reverse, by means of the nomograph.

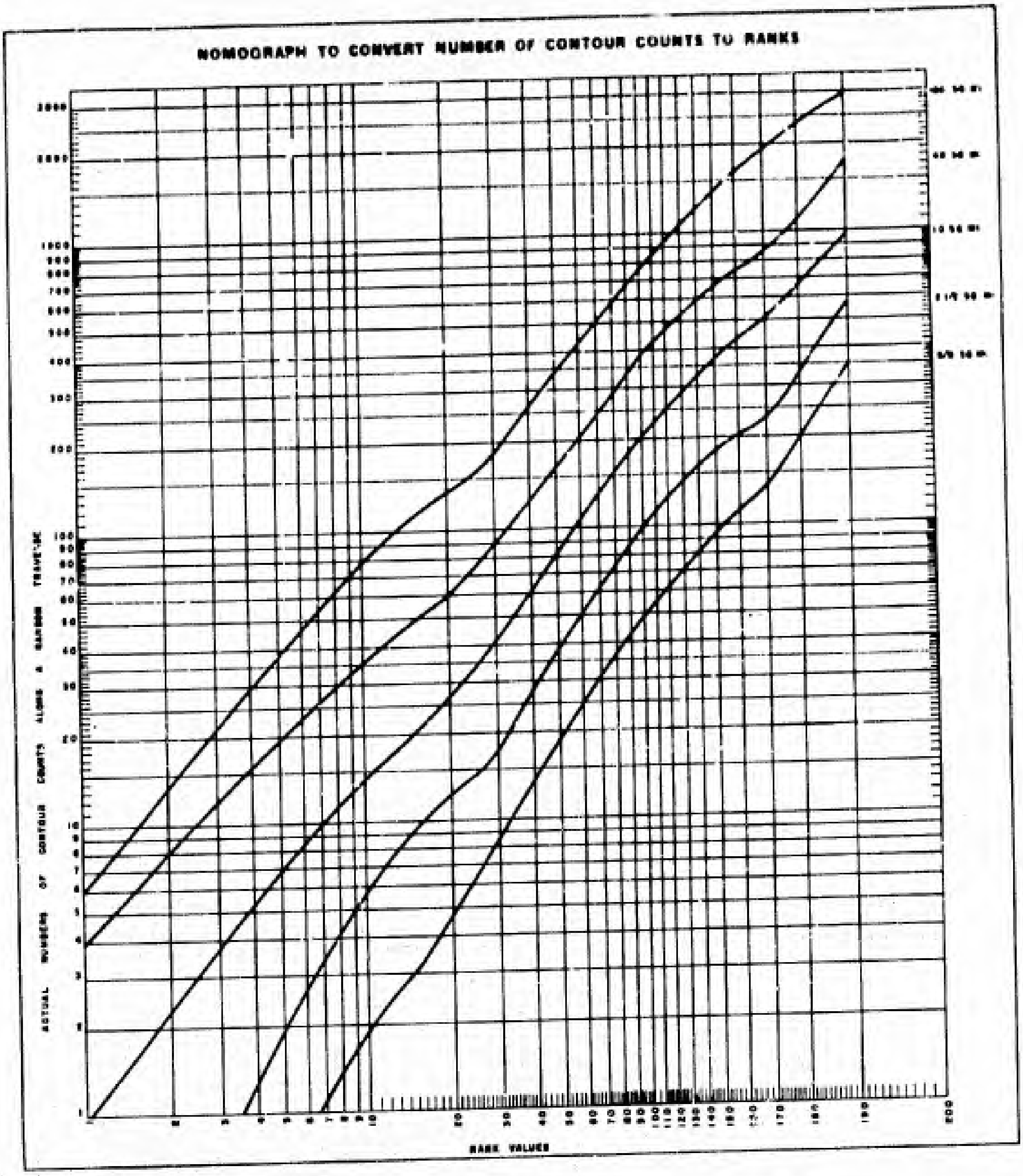


Figure 8. For use in the prediction equations, actual number of contour counts on five sizes of area are converted to rank number of contour counts, or the reverse, by means of the nomograph.

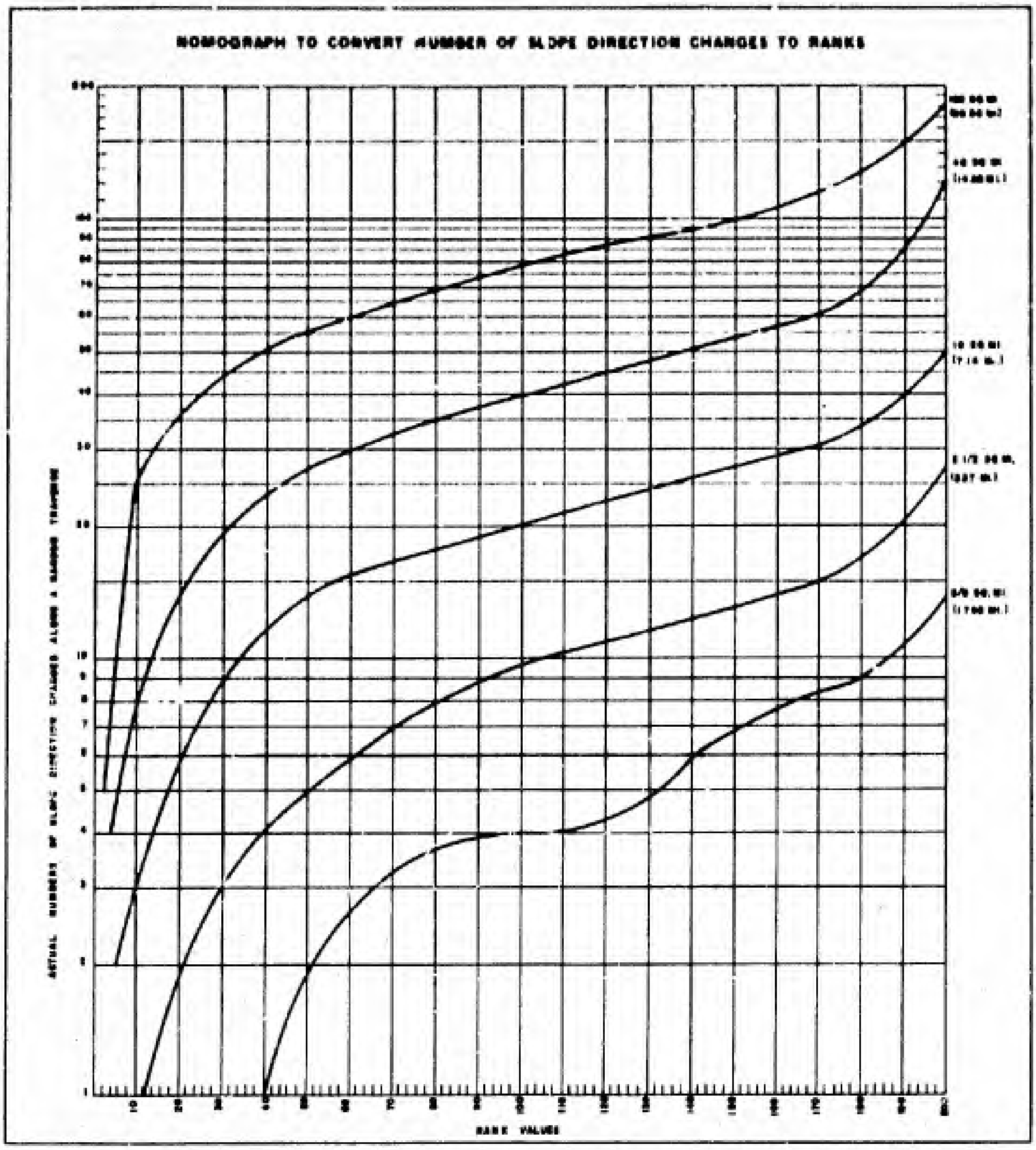


Figure 9. For use in the prediction equations, actual number of slope direction changes on five sizes of area are converted to rank number of slope direction changes, or the reverse, by means of the nomograph.

To demonstrate the greater value of multiple regression over simple regression, data may be taken from the extreme sizes of area (5/8 and 160 square miles), arranged in various combinations, and a series of equations developed. Two test cases were performed to illustrate this technique which indicated those elements that would best improve the original predictions.

The first series of equations predicted relief on 5/8-square-mile samples with information taken from 160-square-mile samples. The three factors, relief, contour counts, and slope direction changes, used independently to predict relief, resulted in the following correlations:

<u>To predict relief on 5/8 square miles (simple regression)</u>:

			Correlation r_r	r_r^2
From	1. Slope direction changes	160 sq miles	.3079	.0949
	2. Relief		.7856	.6172
	3. Contour counts		.8848	.7829

The three factors used in all possible combinations produce the correlations listed below.

<u>To predict relief on 5/8 square miles (multiple regression)</u>:

			Correlation r_r	r_r^2
From	4. Relief and slope direction changes	160 sq miles	.8262	.6827
	5. Contour counts and slope direction changes		.8872	.7871
	6. Contour counts and relief		.8897	.7916
	7. Contour counts, slope direction changes, and relief		.8899	.7919

The r_r's of the simple regression show that slope direction changes on 160 square miles, as a predictor of relief on 5/8 square miles, is very poor; relief as a predictor of relief is good, and "contour counts" is the best predictor of relief. The correlation coefficients of the multiple regression equations indicate that estimates can be improved by adding other factors to the equation. The best estimate of relief, allowing for 79.19% of variance, would result from the equation containing all three factors as predictors. However, this equation contributes no significant improvement over the simple regression equation based on contour counts alone. Although "slope direction changes" by itself is not a good predictor of relief, when added to either of the other two

factors it helps to improve the estimate. (Relief by itself accounts for only 61.72% variance; in combination with slope direction changes, 68.27% of the variance is accounted for.)

The other series of equations predicts contour counts on 5/8-square-mile samples from 160-square-mile information. The system used above is followed:

<u>To predict contour counts on 5/8 square miles (simple regression)</u>:

			Correlation r_r	r_r^2
From	1. Slope direction changes	160 sq miles	.4141	.1718
	2. Relief		.7245	.5249
	3. Contour counts		.8955	.8219

<u>To predict contour counts on 5/8 square miles (multiple regression)</u>:

			Correlation r_r	r_r^2
From	4. Relief and slope direction changes	160 sq miles	.8119	.6591
	5. Relief and contour counts		.8960	.8029
	6. Slope direction changes and contour counts		.8967	.8041
	7. Relief, slope direction changes, and contour counts		.8967	.8041

These examples demonstrate the improvement in some estimates made possible by use of multiple regression equations. They also show that a count of contours along a random traverse (Equation 3) can, by itself, yield nearly as good an estimate when used alone as when used in combination with other data.

Although "contour counts", which may be converted directly into average slope, is one of the most discriminating of terms describing surface geometry, it is, at the same time, the most difficult information to obtain for many parts of the world. Multiple regression equations for predicting contour counts from relief and slope direction changes (Equation 4) on unit areas would have practical use on many foreign maps which show ridge lines, valleys, and spot elevations, but no contours. The principle would be applicable also to the analysis of aerial photographs from which average slope readings are more difficult to obtain than relief values and slope direction changes.

TABLE VIII

MULTIPLE REGRESSION EQUATIONS FOR PREDICTING CONTOUR COUNTS
FROM RELIEF AND SLOPE DIRECTION CHANGES ON UNIT AREAS

AREA	EQUATION	R	R^2
5/8 sq mi	$T = -5.1715 + .8793X_1 + .1722X_2$	.9602	.9218
2 1/2 sq mi	$T = -8.1287 + .8383X_1 + .2426X_2$	.9557	.9136
10 sq mi	$T = -12.3666 + .8061X_1 + .3169X_2$	.9419	.8872
40 sq mi	$T = -15.0938 + .8106X_1 + .3395X_2$	.9261	.8576
160 sq mi	$T = -16.6761 + .8045X_1 + .3614X_2$	.9038	.8168

T = predicted contour counts expressed as a rank value

X_1 = relief expressed as a rank value

X_2 = slope direction changes expressed as a rank value

Equations for predicting contour counts from relief and slope direction changes are presented in Table VIII. When the r_r values of Table VIII are compared with similar cases of the simple regression equations, it is obvious how much the estimates have been improved. For instance, when predicting contour counts on 160 square miles from relief and slope direction changes, r_r is .9038; predicting from relief alone, r_r is only .7245 (Table II).

(3) Performance of the equations

To test the effectiveness of the equations, aside from their theoretical value as indicated by the r_r, rank contour counts on the five unit areas were predicted by using the multiple regression equations presented in Table VIII. The predicted rank countour counts were expressed as an actual number of contour counts which were then converted to a slope tangent by means of the Wentworth equation:(8)

$$\text{Tan.} = \frac{\text{contour counts per mile x contour interval}}{3361}$$

The difference between the predicted and actual slopes was mapped for the two extremes in sample size (5/8 and 160 square miles) in terms of percent slope (Figs. 10 and 11).

In general, the equations perform well in areas of low relief such as the Coastal Plain and Interior Lowland. The results are poorer in areas of high relief such as the Appalachians, Northern Rockies, Cascades, Colorado Plateau, and Ozarks. Although the performance of the equation is not as good for the 160-square-mile samples as for the 5/8-square-mile samples, the same areal pattern of errors occurs.

However, if one were to decide that a prediction error of 4% of slope were tolerable, it could be concluded that the equations used to construct Figures 10 and 11 are adequate for predicting average slope for over 75% of the country. This, of course, assumes that the sample of 200 topographic maps adequately represents the country as a whole.

(4) Sample adequacy

The question might well have arisen in the reader's mind as to the adequacy of this sample of 200 cases to represent the terrain of the United States. If it is representative, then the formulae which have been derived could, in the absence of any better method, be applied to a closer network of samples and maps of different terrain elements produced. If this sample does not adequately represent the surface geometry of the United States, then the findings apply to the sample specifically, and only generally to the rest of the country.

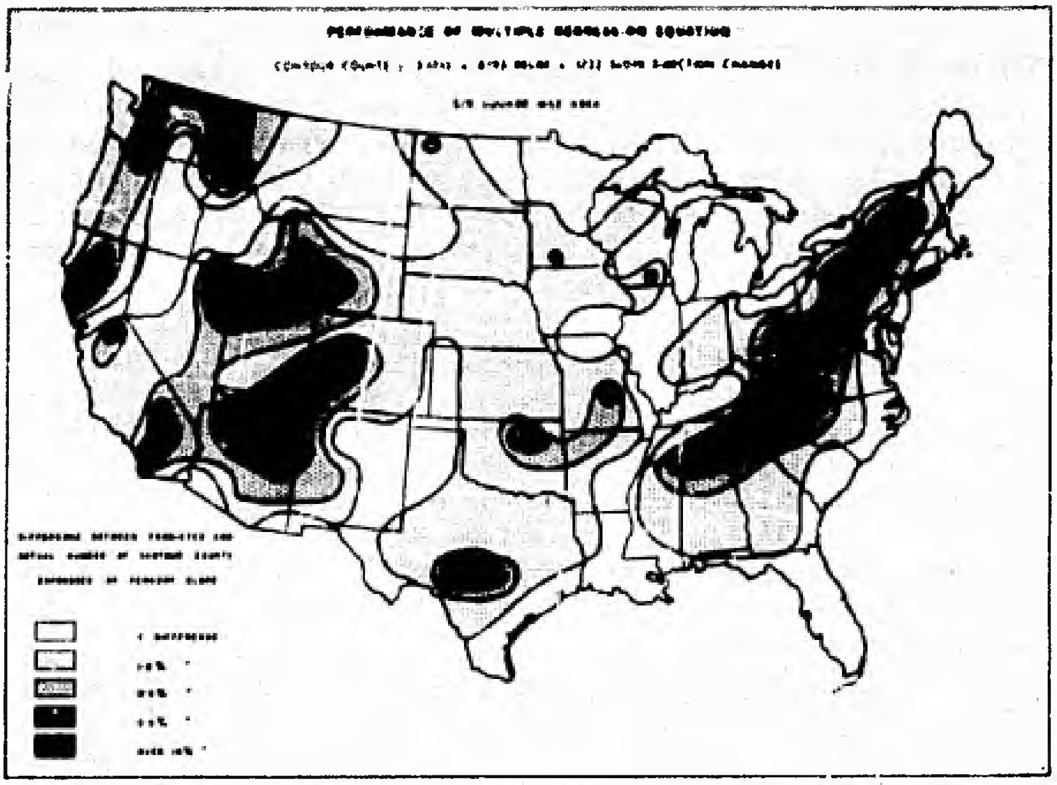


Figure 10. Mapped differences between the actual number of contour counts on 5/8 square mile and the predicted number (converted to % slope), as derived from a multiple regression equation (Table VIII), indicate the areal distribution of the equation's effectiveness.

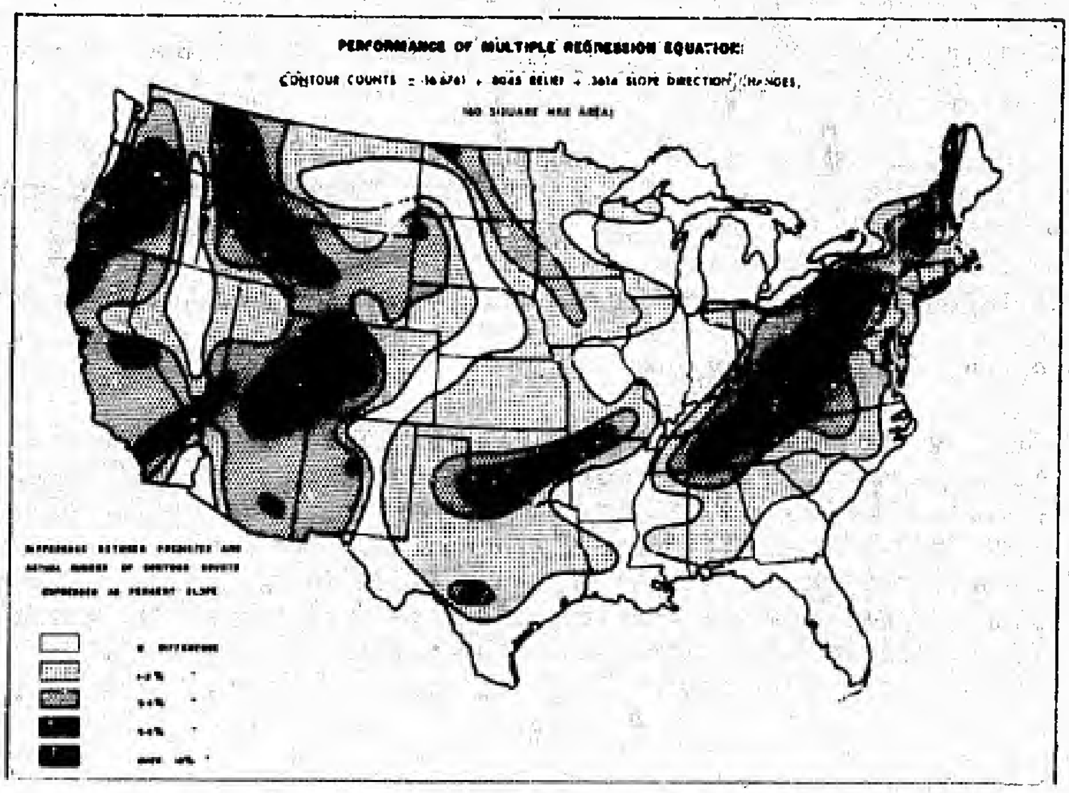


Figure 11. Mapped differences between the actual number of contour counts on 160 square miles and the predicted number (converted to % slope), as derived from a multiple regression equation (Table VIII), indicate the areal distribution of the equation's effectiveness.

It is not planned to give a final and definitive answer to this query since the possibility exists that the sample may be adequate in some respects and inadequate in others (i.e., adequate with regard to average slope on 160 square miles, but not adequate with regard to relief on 5/8 square miles). However, it is possible to obtain a first impression of the sample's adequacy.

It may be safely assumed that a sample satisfactorily represents the population from which it is drawn if it meets these tests: (1) essentially the same means and standard deviations are obtained when the sample has been randomly divided into two equal groups and (2) a regression equation based on one half of the sample shows the same performance characteristics in both halves.

When these tests were performed, it was found that the means and standard deviations of the two sub-groups of the sample and the performance of the prediction equations were similar. (See Appendix.) Thus, it may be said that the sample is adequate in that it represents a statistical distribution of the population of all of the possible 160-square-mile units in the United States. In other words, the nomograph (Fig. 8) showing the ranking of the two hundred 160-square-mile samples according to contour counts (indicating average slope) would show little difference from a chart based on every 160-square-mile area in the country.

The adequacy of this 200-case sample to represent the terrain population of the United States should not be confused with the accuracy of predictions based on it. It has already been noted (Figs. 10 and 11) that predictions of average slope are not reliable in areas of high relief. Better estimates of average slope in these areas may require a closer network of samples and additional data concerning such factors as natural vegetation and bedrock.

5. Conclusions

This report has been written at a stage in terrain research where it is convenient to take stock of the work accomplished to date. Goals for future study, and methods for achieving them, can be formulated on the basis of the findings published in this report.

Again, the orderliness of the earth's surface has been demonstrated, this time by the conversion of terrain data to a rectangular distribution prior to analysis and the presentation of the results on graphs and charts. More details of the pattern and order which exist among landform features will be revealed through extended analysis of these same data. Furthermore, the technique of conversion employed here provides a relatively quick and easy means to manipulate quantified terrain information.

With the correlation and subsequent analysis of six terrain factors, some of the relationships existing among geomorphic elements have been discovered. Moreover, as understanding of these relationships has come about, it has been possible to devise regression equations for predicting various terrain elements where pertinent data may be lacking. A series of regression equations has been formulated for predicting relief and average slope. Maps of the performance of these equations have pointed out those areas not likely to require further study (insofar as relief and average slope are concerned) as well as other areas which will require further study.

The prediction equations presented here apply only to the continental United States. Other areas of the world are also orderly, but their order differs from that of the United States to the same degree as do the structural materials and the physical processes which have shaped them. However, the same techniques of sampling and data analysis used to study the United States can be applied to other regions.

6. Bibliography

1. Davies, G.R. and Yoder, D. Business Statistics, 2d ed, John Wiley and Sons, Inc., New York, 1941, pp 586-589. (Reprinted by permission from Davenport and Ekas, Statistical Methods, 4th ed, John Wiley and Sons, Inc., New York and from G.W. Snedecor, Analysis of Variance, Collegiate Press, Ames, Iowa.)

2. Goldhill, Utah. USGS topographic map, 1928, scale 1:62,500.

3. Horton, R.E. "Erosional Development of Streams and Their Drainage Basins; Hydrophysical Approach to Quantitative Morphology", Bull. of the Geol. Soc. of Am., Vol. 56, pp 275-370, Mar 1945.

4. Messena, New York. USGS topographic map, 1941, scale 1:62,500.

5. Peltier, L.C. "Some Properties of the Average Topographic Slope", abstract, Annals of the AAG, Vol. XLIV, No. 2, Jun 1954, pp 229-230.

6. Playfair, J. Illustrations of the Huttonian Theory of the Earth, Edinburgh, 1802, Reprint, University of Illinois Press, Urbana, Ill., 1956.

7. Strahler, A.N. "Equilibrium Theory of Erosional Slopes Approached by Frequency Distribution Analysis", Am. Jour. of Sci., Vol. 248, Oct 1950, pp 673-696.

8. Wentworth, C.K. "A Simplified Method of Determining the Average Slope of Land Surfaces", Am. Jour. of Sci., Vol. 20, 1930, pg. 186.

9. Wood, W.F. "Increasing the Value of Small Scale Maps for Landform Study by the Use of Statistical Inference", abstract, Annals of the AAG, Vol. 46, No. 2, Jun 1956, pp. 282-283.

10. ----- "The Relationships of Relief", abstract, Annals of the AAG, Vol. 45, No. 3, Sep 1955, pp 306-307.

11. Wood, W.F. and J.B. Snell. The Dispersion of Geomorphic Data around Measures of Central Tendency and Its Application, Research Study Report, EA-8, QM R&D Center, Natick, Mass., Mar. 1957.

APPENDIX

PROOF OF SAMPLE ADEQUACY

The sample was divided into two groups of 100 cases each by allowing disinterested co-workers to shuffle the data cards and divide them into two piles, labeled Groups I and II. The means and standard deviations of these two groups, along with those of the total sample, are as follows:

		Group I	Group II	Total Population
Contour counts	Mean number*	542.16	531.36	536.76
	Standard deviation	559.67	525.10	538.07
Average slope tangent	Mean	.1132	.1109	.1121
	Standard deviation	.1169	.1096	.1121

*Mean number of 20-foot contour counts along 2 diameters of a 160-square-mile circle.

The data of Group I transformed into a rectangular distribution by the ranking method already demonstrated, produced the following regression equation for predicting contour counts on 160 square miles from slope direction changes and relief on 160 square miles: $T = -9.0955 + .8137 X_1 + .3664 X_2$ where "-9.0955" is a constant, "X_1" is rank value of relief, and "X_2" is rank value of slope direction changes. (It will be seen that this equation is nearly identical to the equation developed for the same purpose from the 200-case sample: $T = -16.6761 + .8045 X_1 + .3614 X_2$ (Table VIII). The X_1 and X_2 values of the new equation differ very little from the first. The value for "a" is about half, since 100 rather than 200 samples are used.) When this equation is used to predict the contour counts which ought to be present, and the prediction compared to the actual number of contour counts, a further comparison of the two halves of the sample is possible (Table IX).

Table IX is a tabulation of errors by magnitude, whether plus or minus. Because of the relatively wide range of errors, their comparison is best made by the use of a cumulative curve chart of errors. This is done in Figure 12, based on the data from Table IX. Except at the ends of the distribution, the distribution of errors is similar in both groups. Underestimates greater than 25% slope and overestimates greater than 5% involve only 10 of the 200 cases. Either the symbol "+", indicating

errors in Group I, or the symbol "o", indicating errors in Group II, could be used as guides for drawing a smooth line to represent the distribution of errors in either or both samples. As would be expected, the greatest errors occurred in the same geographical location as did the errors mapped in Figure 11. Both sample Groups I and II exhibit the same distribution of errors and are biased toward underprediction.

TABLE IX
DISTRIBUTION OF AVERAGE SLOPE PREDICTION ERRORS
Two Groups of 100 Cases

Error Percent Slope	Group I: Number of Cases	Group I: Cumulative Number of Cases	Group II: Number of Cases	Group II: Cumulative Number of Cases
-36	1	1		
25	1	2	1	1
24			1	2
23			1	3
22	1	3		
19	1	4	1	4
17	1	5		
15			1	5
13	2	7	2	7
12			1	8
11			1	9
10	1	8		
9			1	10
8	3	11	3	13
7			2	15
6	4	15		
5	3	18	1	16
4	3	21	4	20
3	2	23	1	21
2	6	29	7	28
1	14	43	10	38
0	14	57	22	60
+ 1	11	68	11	71
2	14	82	8	79
3	4	86	6	85
4	5	91	6	91
5	4	95	4	95
6	1	96		
7	2	98		
9			1	96
11			1	97
13	2	100		
15			1	98
19			1	99
50			1	100
Mean error		3.99		4.43
Standard deviation of error		5.63		7.10

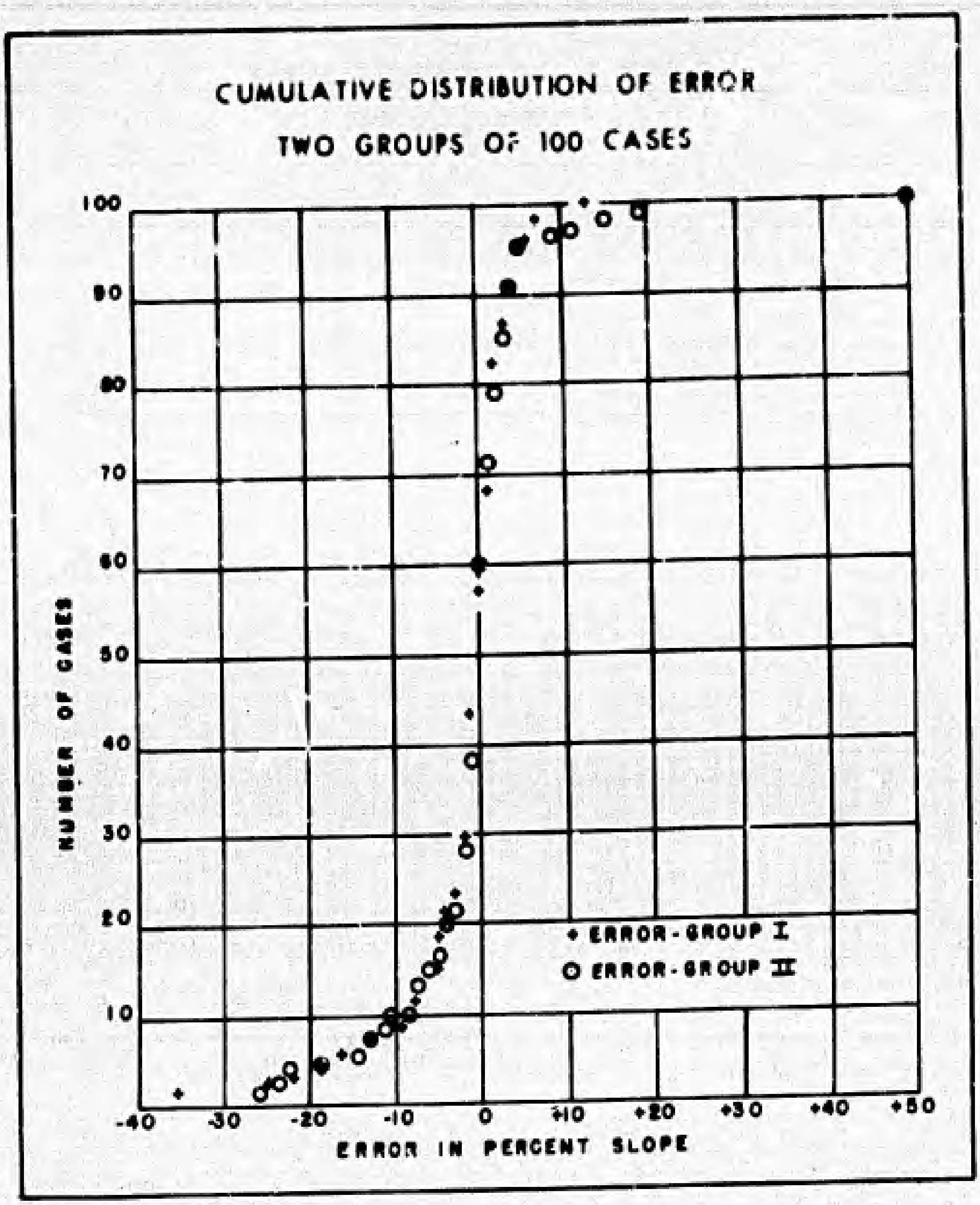


Figure 12. Sample adequacy is, in part, proved by the performance of a regression equation based on half of the sample, which is used to predict average slope for both halves of the sample.